Workshop Report:

Improving the Quality and Reuse of Cybersecurity Datasets, Software, and Other Research Artifacts

March 2023

Report for the workshop held virtually on September 15, 2022

Table of Contents

Preface **3**
Workshop Organizers **4**
Executive Summary **5**
Introduction **8**
Workshop Goal 8
Workshop Participation 8
Workshop Agenda 9
Discussion Sessions **10**
Reflections on Artifact Evaluation 10
Artifact Repositories, Indices, and Challenges 12
Artifact Evaluation Initiatives at Conferences 14
Artifact Quality and Metadata Standards 16
Beyond Datasets and Code 17
Next Steps - What Can We Do as a Community? 18
Artifact Survey **20**
Key Findings **23**
Conclusions and Next Steps **24**
Acknowledgements **26**
References **27**
Appendices **29**
Appendix A: List of Acronyms 29
Appendix B: Workshop Invitation 30
Appendix C: Workshop Agenda 33
Appendix D: Workshop Presentations 36
Appendix E: Summary of Artifact Evaluation Initiatives at Cybersecurity Conferences 37
Appendix F: Surveying Sharing and Reuse of Computational Artifacts 41

Preface

This report summarizes the presentations, discussions, and key findings from the workshop, “Datasets, Software, and Other Research Artifacts,” held virtually September 16, 2022. Workshop participants attended virtually via Zoom. The workshop was attended by 32 people from 18 organizations across the United States as well as Austria and Italy. It served as a forum for learning and understanding the issues and challenges researchers face in producing, evaluating, and reusing high-quality cybersecurity research artifacts, including datasets and software, as well as experiment designs, execution, and results. The workshop benefited from the degree of openness and interaction displayed by the participants while discussing the challenges and opportunities for improving the quality and result of cybersecurity research artifacts. By openly sharing their experiences and knowledge, insights were gained which are documented in this report and which should provide value to all the participants and their respective organizations. Furthermore, the workshop was the first in what will hopefully be an ongoing series of discussions and interactions and the start of a new community centered around improving the sharing and reuse of cybersecurity research artifacts.

Workshop Organizers

David Balenson
Senior Computer Scientist, Networking and Cybersecurity Division, USC Information Sciences Institute
balenson@isi.edu

Terry Benzel
Director of the Networking and Cybersecurity Division, USC Information Sciences Institute
tbenzel@isi.edu

Eric Eide
Research Associate Professor, School of Computing, University of Utah
eeide@cs.utah.edu

David Emmerich
Principal Cyber-Physical Range Architect, Information Trust Institute, University of Illinois Urbana-Champaign
davidpe@illinois.edu

David Johnson
Research Associate and Systems Software Engineer, School of Computing, University of Utah
johnsond@cs.utah.edu

Jelena Mirkovic
Research Associate Professor and Research Team Leader, Networking and Cybersecurity Division, USC Information Sciences Institute
mirkovic@isi.edu

Laura Tinnel
Senior Computer Scientist, Computer Science Laboratory, SRI International
laura.tinnel@sri.com

Executive Summary

A one-day workshop entitled “Improving the Quality and Reuse of Cybersecurity Datasets, Software, and Other Research Artifacts,” was held virtually on September 15, 2022. Additionally, we constructed and distributed an artifact survey to gain input from a wider community of researchers. This report summarizes the presentations, discussions, and key findings from the workshop as well as the results of the survey.

Researchers in experimental cybersecurity are increasingly sharing the code, data, and other artifacts associated with their studies. This trend is encouraged and rewarded by conferences and journals through practices such as artifact evaluation and badging. While these trends in sharing artifacts are promising, the cybersecurity community is still far from an ecosystem in which artifacts are FAIR: findable, accessible, interoperable, and reusable. The lack of established sharing and reuse standards and best practices results in artifacts that are often difficult to find and reuse; these artifacts may also be incomplete or of low-quality.

The overall goal of the workshop was to explore the issues and challenges researchers face in producing, evaluating, and reusing high-quality cybersecurity research artifacts, including datasets and software, as well as experiment designs, execution, and results. The workshop brought together a broad group of 32 researchers, professors, graduate students, and others from 18 organizations, mostly across academia, but also from industry and government, to discuss their experiences and interests in producing, evaluating, and reusing experimental research artifacts.

During the workshop, attendees participated in a series of interactive sessions to discuss artifact evaluation; artifact repositories, indices, and challenges; artifact evaluation processes at conferences; artifact quality and metadata standards; going beyond datasets and code; and next steps - what can we do as a community? Additionally, we conducted an artifact survey to gain input from a wider community of researchers beyond the workshop participants.

Together, the workshop and survey produced key findings around artifact findability, artifact quality and usability, artifact maintenance, incentives and funding, and the full research lifecycle:

- **Artifact findability:** While there are currently multiple channels where researchers share their artifacts (e.g., Github repositories, lab and personal Web pages, Zenodo) there are issues and challenges. There is a need to improve the ability to find artifacts, by providing locations where artifacts can be permanently archived and ways to index shared artifacts across multiple platforms and provide a common repository for this index.
- **Artifact quality and reusability:** There is a great need to improve artifact quality and usability, including better code packaging, better metadata, and better documentation for artifacts. It is clear that the community needs to create standards around packaging, metadata, and documentation and needs to teach young researchers how to produce artifacts that meet these standards.

- **Artifact maintenance:** There is a significant burden in maintaining artifacts with an unclear payoff for authors. Moreover, there are no established practices to pass artifacts on to others for continued maintenance when authors graduate or move on to new projects. Even though the problem is complex, the community needs to find solutions to these challenges, understanding that only partial solutions may be possible.
- **Incentives and funding:** Artifact authors and evaluators need more recognition and perceived value for their work. Intrinsic incentives for artifact sharing and reuse should exist and work in concert with promotion and tenure pathways, career pathways, and funding. Authors need more funding to produce higher-quality artifacts to perform the additional work.
- **Support for the full research lifecycle:** Artifact support is needed to capture and share all parts of the research lifecycle. Required support extends beyond datasets and code to include research methods, such as the hypothesis, workflow, experiment design, experimentation environment, assumptions, limitations, etc.

The workshop discussions and survey results also pointed toward a number of directions and immediate next steps for the community.

- **Standardize requirements and processes around artifact sharing:** Frequent discussion topics were poor artifact quality, artifacts becoming obsolete over time, and artifacts needing to be portable. The community can begin to address these issues by standardizing requirements and processes around artifact sharing and creating educational materials to teach young researchers how to produce compliant artifacts.
- **Create incentives for authors and evaluators:** Today, authors must invest significant effort to prepare, release, and maintain an artifact associated with an existing publication. Similarly, evaluators spend a lot of time assessing artifacts based on some set of evaluation criteria. The community needs to identify and support professional and (perhaps) monetary incentives for authors and evaluators. Some discussions touched on the possibility of requiring artifacts to be published when a research publication is accepted. Recently many venues have started encouraging artifact publication. A few venues require artifact submission (e.g., ACM CCS, Journal of Systems Research).
- **Build a culture of sharing and reuse:** Conferences and journals have an opportunity to help build a culture of sharing and reuse in the research community through continued calls for artifact publication and requirements that the research described in paper submissions be evaluated against published artifacts. Further, using community guidelines, tutorials, university classes, and software packages to help teach young researchers how to share and reuse artifacts will help build a culture of sharing and reuse.
- **Provide funding for artifacts sharing:** Funding agencies should consider supplemental funding vehicles to support artifact-sharing efforts and include precise requirements for measuring artifact quality. Funding vehicles should not be overly competitive or require significant funding.

- **Provide research infrastructure for artifacts sharing**: The community would benefit significantly from robust online catalogs of research artifacts, such as SEARCCH [SEARCH-HUB], with rich metadata and information about artifact quality and applicability. The community would also benefit from an experimentation environment with full support for the research lifecycle, including artifact packaging for sharing and reuse. Such infrastructure could be used by artifact evaluation committees to offer evaluation services directly on the infrastructure. Once an evaluation completes, its records, evaluator notes, and packaged experiments could remain housed on the infrastructure to be easily reused by others in the same environment.

The robust workshop and survey participation, wide-ranging perspectives, and productive discussion sessions and comments demonstrated a strong need for a coordinated effort to improve the quality and reuse of cybersecurity research artifacts. The workshop was the first in what will hopefully be an ongoing series of discussions and interactions and the start of a new, vibrant community centered around improving the quality and reuse of cybersecurity research artifacts.

Introduction

Researchers in experimental cybersecurity are increasingly sharing code, data, and other artifacts associated with their studies. This trend is encouraged and rewarded by conferences and journals through practices such as artifact evaluation and badging. While these trends in sharing artifacts are promising, the cybersecurity community is still far from an ecosystem in which artifacts are FAIR: findable, accessible, interoperable, and reusable [FAIR, WILKINSON, BALENSON]. The lack of established standards and best practices for sharing and reuse results in artifacts that are often difficult to find and reuse; in addition, the lack of community standards results in artifacts that may be incomplete or of low-quality.

Workshop Goal

The overall goal of the workshop was to explore the issues and challenges researchers face in producing, evaluating, and reusing high-quality cybersecurity research artifacts, including datasets and software, as well as experiment designs, execution, and results.

Workshop Participation

A broad set of researchers who were producing, evaluating, or reusing cybersecurity research artifacts were invited to participate in the workshop. A copy of the invitation email is contained in Appendix B. The invitation was sent to over 100 people in the community.

The workshop was attended by a broad group of 32 researchers, scientists, and engineers from 18 organizations across academia to discuss their experiences and interests in producing, evaluating, and reusing experimental research artifacts.

- Boston University
- Colgate University
- Colorado State University
- NIST
- Rochester Institute of Technology
- Sandia National Laboratories
- SRI International
- Texas A&M University
- TU Wien
- UCSD/CAIDA
- University of Georgia and GATech
- University of Illinois
- University of Memphis
- University of Padova
- University of Utah
- University of Vermont
- USC Information Sciences Institute
- Virginia Tech University

Workshop Agenda

The workshop featured a series of interactive sessions in which participants discussed the following topics:

- Artifact evaluation;
- Artifact repositories, indices, and challenges;
- Artifact evaluation processes at conferences;
- Artifact quality and metadata standards; going beyond datasets and code; and
- Next steps - what can we do as a community?

A copy of the complete workshop agenda is in Appendix C.

Each session started with a short presentation meant to catalyze the discussion. Links to the slides from the sessions are listed in Appendix D.

Discussion Sessions

The following subsections summarize each of the interactive discussion sessions, including the initial presentation used to seed the discussion and key points and highlights from the subsequent discussion.

Reflections on Artifact Evaluation

Eric Eide (University of Utah) facilitated this session, which introduced the topic of artifact evaluations. Prof. Eide used the following description and questions to stimulate the conversation.

> In the past decade, artifact evaluation has become a feature of many conferences in cybersecurity and other areas of computer science. Widespread practice is for artifacts to be solicited from the authors of accepted papers. Papers receive badges if their corresponding artifacts are available, functional, reusable, and/or useful for reproducing the papers' main results.
>
> - How has the adoption of artifact evaluation affected your practices and/or expectations of published research?
> - How well is the current adoption and practice of artifact evaluation serving the community?
> - What has the community learned from the current practice of artifact evaluation?
> - In what ways should current practice evolve to better serve the research community?

We briefly summarize the discussion during this session.

Prof. Eide started with a brief presentation that reviewed the origins and evolution of artifact evaluation within the computer science and cybersecurity research communities. His presentation offered quotes from several relevant publications and websites and appropriate citations.

The early advocates of artifact evaluation at computer science conferences designed the process with multiple goals in mind. A principal objective was to encourage repeatability, i.e., the ability of a researcher to attempt to validate the scientific claims of a published paper by repeating the experiments described in that paper. Other goals of artifact evaluation included:

1. Improving the quality of research publications, e.g., by encouraging authors to make more specific claims;
2. Encouraging good practices, such as automation and documentation, in the development of research software and datasets;
3. Rewarding authors who create high-quality research software and datasets through the awarding of “badges” to authors; and

4. Enabling derivative research efforts based on previously published work. While common practice is for artifact evaluation to be separate from paper review, the original advocates believed that artifact evaluation should eventually play a role in reviewing works submitted for publication.

Artifact evaluation is now commonplace in computer science conferences and journals, and professional societies such as ACM have adopted standard guidelines for artifact evaluation and badges [ACM-ARTIFACTS]. The push for widely available artifacts has led to other evolutions in the research community. For example, conferences' calls for papers increasingly identify replicability as a criterion for evaluating submissions. Conference committees are also increasingly welcoming reproducibility studies (i.e., attempts to validate previously published work). University professors are enhancing student research training by assigning students to study available artifacts and utilize them to (attempt to) repeat the experiments in published works.

Following his stage-setting presentation, Prof. Eide led the workshop participants to discuss the current practice of artifact evaluation in the computer science and cybersecurity communities, centered on the four questions listed above. We summarize the main points of the discussion.

Value of artifact evaluation: For science, artifact evaluation serves to confirm or refute prior findings or measure the sensitivity of research findings to new environments (small changes may significantly impact results) [MYTKOWICZ]. For authors, the evaluation and reuse of their artifacts may bring recognition in the field and improve their artifact quality. For conferences, artifact evaluation gives credibility to the companion paper (which introduces the artifact) and enhances the overall scientific rigor of the forum. For education, instructors can use evaluated artifacts in classes to demonstrate a principle and be assured they will work as advertised.

Challenges for artifact evaluation: Authors may feel that their artifact is not ready to be evaluated and thus may avoid any assessment. A big obstacle seems to be the lack of students' skills in software development, which leads to poor artifact quality. Faculty must teach students how to produce artifacts that are easy to maintain, reuse and evaluate. Another big obstacle is the low payoff for the authors. Authors receive recognition mostly from publications; artifacts do not bring comparable recognition - this discourages the submission of artifacts for evaluation. Another challenge is that one may want to publish an artifact first (e.g., a dataset) before using it in a paper. It is unclear where to publish artifacts and receive recognition for this. Yet another challenge is recruiting a skilled, motivated set of evaluators. They get almost no recognition, so their motivation to participate is low.

Should artifacts be required for publications: Different venues have different artifact requirements. Some venues require artifacts to be at least submitted for evaluation, while others strongly encourage, but do not require submission, and some still have no artifacts submission and evaluation process. Any action from the venue that promotes artifact evaluation seems to help (i.e., strong encouragement still increases the number of artifact submissions).

Encouraging authors of accepted papers to submit artifacts and providing them additional time to do so may improve submission rates. Venues should clearly state that artifacts do not need to be perfect for submission. Instead, the artifact evaluation (AE) committee works with authors in iterations to improve artifact quality, and everyone wins.

Artifact Repositories, Indices, and Challenges

This discussion session introduced existing artifact repositories and indices and some of the challenges with such infrastructure. Laura Tinnel (SRI International) facilitated the session. Ms. Tinnel gave the group the following description and questions to stimulate the conversation.

> Conferences, workshops, and researchers use several standard paper repositories, including the ACM Digital Library, IEEE Xplore digital library, Elsevier Science Direct, Springer, USENIX, Zenodo, and arXiv.org. Standard data and code repositories also exist and are used by researchers. Examples include GitLab, GitHub, AWS CodeArtifact, Bitbucket, Zenodo, Microsoft Azure (for artifacts), Google Code (for artifacts), SourceForge, and Codebase. Finally, several indices and catalogs allow researchers to store and search for research artifacts in different fields. For example, Google Scholar, ResearchGate, Google Dataset, Mendeley, and OpenAire Explore all catalog papers, data, and code for general scientific research. FindResearch.org catalogs computer science artifacts, Papers with Code catalogs machine learning artifacts, and SEARCCH [SEARCCH, SEARCCH-HUB] catalogs cybersecurity research artifacts. Ms. Tinnel also noted that non-research-focused hubs are popping up to support DevOps, such as ArtifactHub, which supports cloud computing.
>
> SEARCCH is a social cybersecurity artifacts catalog developed by the workshop organizers under an NSF CCRI grant. The SEARCCH catalog focuses on code and data, although publications are also supported, and adding associated papers is encouraged to enable rich relationship knowledge. The catalog does not store artifacts, just catalog entries and social interactions. Anyone can search for artifacts by author, organization, type, badge, owner, description, etc. Users can create accounts and profiles to take ownership of artifacts, identify favorites, and review and rate artifacts. Semi-automated importers are available to add artifacts to the catalog. Existing importers include support for git-based repositories, ACM Digital Library, IEEE Xplore, USENIX Security, NDSS, Papers With Code, Zenodo, and arXiv. Given a target URL, the importer parses the webpage and objects and suggests values for catalog fields. Curators and authors can edit imported artifact catalog entries as needed before publishing. Curators and authors may add artifact entries through a manual process using a catalog entry editor. The catalog also allows an author to take ownership of previously imported artifact catalog entries and edit the entry to improve or correct information. Prior catalog entry versions are preserved, much like the Wikipedia model.

Following her introduction, Ms. Tinnel led the workshop participants in a discussion about how to increase useful sharing and reuse based on the following questions:

- How do we balance the need for sharing with "useful" sharing?
    - Will the adoption of artifact packaging and metadata standards discourage sharing? (e.g., dataset metadata standards that aren't required but, if used, increase your content's searchability in their engine.)
    - What tools would help with sharing? E.g., Standards-based packaging wizard that outputs standard directory structure(s), keywords, metadata?
- How much effort is reasonable for initially adding artifacts to indices?
    - Hypothesis: less time => greater use, adoption
    - Assistance vs. automation
    - Would automated crawling be better / worse than manual indexing?
- Is crowdsourcing knowledge about artifact use (over time) helpful to researchers?
    - Author ownership/verification of artifact information
    - Community discussion
    - How to incentivize? Badges? Other?
- What additional catalog features would be the most useful to the cybersecurity research community?
    - Research catalog notebook?
        - Next step to "favorites" -- the ability to annotate, sort, and organize favorites
        - Ability to add things found outside of the SEARCCH catalog and have them automatically imported and added to the notebook.

We summarize the discussion during this session.

Findability of artifacts: Researchers have multiple places where they may search for artifacts. Often, the value lies in identifying and exposing relationships between artifacts (e.g., an artifact from a paper or an artifact that is an improved version of another artifact), especially if searching for something useful in a big area. New catalogs need to acquire a critical mass of users and artifacts to become well-used and motivate authors to put their artifacts directly into the catalog. Catalogs could also work with artifact evaluation committees to have the author submit their artifact through the catalog.

Estimating artifact usability: Many artifacts do not have standardized metadata, which makes assessing usability challenging for researchers. The community could improve this process by developing guidelines and tutorials around artifact metadata. There may also be ways to capture usability in the artifact catalog via user reviews. However, positive reinforcement usually works better than negative reinforcement, especially when made public. This points to the need for a reward-based system.

Initial and long-term artifact maintenance: In academic environments, graduate students are usually responsible for packaging and maintaining artifacts. When a graduate student leaves, a

question arises over who will take on the longer-term maintenance of a usable artifact. This creates a lot of additional work for faculty. If no one takes ownership and responsibility for the departing student's artifacts, these artifacts may become stale and "age out."

Artifact Evaluation Initiatives at Conferences

This session discussed conference artifacts evaluation initiatives and processes at events such as ACSAC, USENIX Security, and PETS. David Balenson (USC Information Sciences Institute) facilitated this session. Mr. Balenson presented the following questions to stimulate the conversation:

- What works well with the current artifact evaluation initiatives?
- What guidance is provided to artifact submitters? To artifact evaluators?
- How could the current processes be improved?
- How could the community capture the knowledge gained during the artifacts evaluation process?

The Annual Computer Security Applications Conference (ACSAC) was the first cybersecurity conference to institute an artifact evaluation initiative in 2017 [ACSAC-AE17]. The initiative continues today [ACSAC-AE22]. To help support research result reproducibility, the conference encourages authors of accepted papers to submit the software they develop and datasets they use to perform their research and make them publicly available to the entire community. An Artifacts Evaluation Committee reviews artifacts to ensure they are documented, consistent, complete, and exercisable. Papers associated with accepted artifacts receive an ACM Functional or Reusable badge [ACM-ARTIFACTS] and are included on the conference website. There is also a Distinguished Paper Award reserved for this group.

In 2022, ACSAC further instituted a Cybersecurity Artifacts Competition and Impact Award, which is open to cybersecurity artifacts previously published in all peer-reviewed venues (conferences and journals) in academic and industry, not just ACSAC [ACSAC-COMP]. Eligible submission must be: related to applied security or privacy; consist of open-source software, open or reusable dataset, free online research service, open testbed, and/or user study materials; be part of published academic or industry research; published before 2020; and had a meaningful impact with benefits for the security research community.

USENIX Security also instituted a Call for Artifacts, starting in 2020 and continuing to today [USENIX-AE22]. The symposium encourages authors of (conditionally) accepted papers to submit artifacts that are reviewed by an AEC. Papers with accepted artifacts receive ACM Artifacts Available, Artifacts Functional, or Results Reproduced badge [ACM-ARTIFACTS]. To help facilitate the packaging and review of artifacts, the symposium provides extensive suggestions and guidelines [USENIX-AE22, USENIX-AE22-GUIDE].

In 2022, the ACM Conference on Computer and Communications Security (CCS) started requiring that “submissions whose claimed contributions rely on artifacts (e.g., code, models,

data sets) make these accessible to the reviewers unless there are good reasons not to, in which case these reasons must be mentioned in the submission." [CCS-AE22].

There are numerous sources for artifact guidelines and best practices, including the USENIX Security CFP [USENIX-AE22] and Appendix Guidelines [USENIX-AE22-GUIDE] and other sources [BAROWY, BELLER, PADHYE, XU].

ACSAC and ACM CCS use the ACM Artifact Review and Badging system [ACM-ARTIFACTS]. The system has three levels of badges: Artifacts Evaluated, Artifacts Available, and Results Evaluated. Artifacts Evaluated means that artifacts have been evaluated as either Functional (documented, consistent, complete, and exercisable) or Reusable (functional plus documented and structured to facilitate reuse and repurposing.) Artifacts Available means that artifacts have been made permanently available but have not been evaluated. Results Evaluated is for papers where the main results have been replicated by a person or team other than the author.

The Security Research Artifacts website links to resources and results around artifact evaluation for security conferences and workshops, including ACSAC, and USENIX Security, among others [SECARTIFACTS].

Appendix E summarizes the artifact evaluation initiatives at ACSAC, USENIX Security, and ACM CCS through 2023.

We summarize the discussion from this session.

Success stories: From ACSAC - the process has evolved and improved over the years. Nowadays, artifact submission and evaluation discussions happen through anonymous channels, and around 30% of papers submit artifacts. There are also extensive guidelines for authors and evaluators. From PETS - multiple evaluators, more valuable and actionable feedback. The artifact competition at ACSAC 2022 [ACSAC-COMP] seems quite popular and may help promote ideas around artifact sharing.

Challenges: Someone needs to follow up and mediate between authors and evaluators (usually students) to ensure that the evaluation starts on time and that discussion is progressing and is civil. AE chairs or student mentors typically do this. Also, some artifacts need specialized hardware, and authors must make such hardware available to evaluators. This can sometimes be challenging. Further, some authors do not understand the scope of an artifact - multiple artifacts with limited scope are more useful than one monolithic artifact. Examples and guidelines here would help. Another challenge is that authors fear their code is too messy or incomplete to share. AE committees should clarify that this is expected and all code is useful to share. Yet another challenge is that it is difficult to reproduce research results. It is much easier to verify that the artifact is functional.

Artifact Quality and Metadata Standards

This session focused on the issues of artifact quality. While many artifacts are shared today, a few are reused due to artifact quality issues. While no specific definition exists around what "artifact quality" means, the community generally understands that quality artifacts are easy to deploy in various settings, easy to run and ensure it is running correctly, and easy to configure and customize. Documentation plays a significant role in artifact quality. Sadly, many artifacts shared today have poor documentation and are not readily usable. One way to address usability and portability could be to package artifacts in a Docker or VM image. This approach would work for artifacts that focus on computation and are relatively simple (single application) but would prove challenging for other, more complex artifacts. In addition to packaging, another open problem is how to support artifact owners in developing high-quality artifacts. The required effort is often unpaid, unseen, and time-consuming. A related issue is the community-based evaluation of artifact quality. Today, high-quality artifacts are easy to identify because they are heavily used. However, owners of other artifacts do not have feedback from the community if their artifacts are uninteresting or if they are needed but have usability issues. Ideally, the community would reward young researchers for evaluating an artifact and posting its review somewhere so the owner can take reparative action.

Here we summarize the discussion for this session.

Measuring artifact quality: For code, high quality means that code runs easily and reliably and that the documentation is good (comprehensive, detailed). For datasets, high quality means that the dataset is well documented, limitations are well documented, and the dataset is representative, large, and hopefully recent. There may not be an objective way to measure quality, so the community may have to rely on subjective feedback from users who tried to use the artifact.

Challenges around producing high-quality artifacts: Authors may be unsure if anyone will use the artifact. They may prefer to publish an imperfect version and then improve it based on questions from potential users. Authors are often students and may lack software development skills. They need to be taught those skills (e.g., Docker, VM, unit testing) and know from the very onset of research that there is an expectation that an artifact will be released. Authors may also feel that time spent on artifact development is wasted because paper publications are the only thing that matters for graduation or promotion. Another challenge is maintenance. Software dependencies evolve, and it is hard to maintain old artifacts. This challenge is compounded by artifact authors graduating and leaving. This challenge may be addressed by sharing Docker and VM images. Also, new users may be able to restore and improve the code. We need some way to recognize this kind of contribution, similar to the co-authorship of the artifact. Another challenge is that artifact evaluation does not bring recognition to evaluators and does not capture lessons learned from evaluation. We could improve this by documenting the anonymized evaluation process and recognizing evaluators by name.

Funding challenges: Many issues around artifact production and quality can be addressed with sufficient funding. Today this is lacking. We need mechanisms to fund artifact release as an add-on to funded research. This funding should not be significant (e.g., NSF Transition to Practice supplements (TTP)) or difficult to obtain. Instead, it should be as easy as NSF Research Experiences for Undergraduates (REU) supplements and provide a few months' funding for a student or two to produce a high-quality artifact from already published research. Funded artifacts must be checked for quality before being accepted, i.e., quality should be the requirement when funding the effort.

Beyond Datasets and Code

This session discussed sharing other experimental research artifacts. conference artifacts. Terry Benzel (USC Information Sciences Institute) facilitated the session. She noted that there has been a rapid increase in data and code sharing through a wide range of processes, infrastructure, and reward mechanisms, but most shared artifacts are code and data tied to a publication. She asked what other experimental artifacts of research that should be shared and how much information is needed about research infrastructure to share an "experiment".

Ms. Benzel posed the following questions to start the discussion:

- How do we capture and share experiment designs, execution, and results?
- What is an experiment design - scripts, topologies, execution steps?
- What about research infrastructure dependencies?
- How do we share processes, steps, and life-cycles meaningfully?

We summarize the discussion for this session.

Research methods: How do we capture and share research methods, such as hypothesis, workflow, experiment design, and experimentation environment? USC-ISI's work on DIstributed Experiment Workflows (DEW) provides a way to encode experiment behavior [MIRKOVIC]. Some aspects of an environment may not be captured because they are not apparent to the researcher. One approach is to share the environment - release an artifact on the shared resource (e.g., testbed), and others will reuse it there. Another way is to use Docker or VM images, which ultimately can export and freeze the most relevant environment settings. Sharing workflows is challenging because experimentation may be manual and happen over long periods; users forget what they tried and what ultimately worked. We need automated ways to capture this data since it is too much to rely solely on user training and cognition. Ideally, all experimentation would be automated at a high level, but this happens very late in a project.

Well-thought-out and planned experiments with multiple trials are prime candidates for automation. A transition from interactive experimentation to automated experimentation is needed. Experiment automation tools need to capture and re-execute experiment steps. Also,

experimenters need to be able to move back and forth between interactive and automated modalities. Experiment steps should be easily intertwined across the two modalities.

Another point around sharing research methods is that students may not grasp the importance of doing this since many publications only focus on results. What is needed are classes, materials, workshops, and conferences that communicate the importance of research methods and show the process of trial and error and learning from mistakes. A counterargument is that some paper submissions may be rejected if they provide too many details about research methods since this gives reviewers something to critique.

Describing experiment design may be domain-specific. One needs to define the assumptions, the limitations, how representative the experiment design is, and the sources of variability. Trends in security and privacy are going from more specified, closed experiments to mixed-methods experiments that interact with the real world. Repeatability is difficult to ensure for such mixed-methods experiments. Another aspect is whether the researchers conducted enough experiment trials to arrive at reliable conclusions and discount the effect of randomness. Also, parameter space needs to be sufficiently explored in an experiment. The community needs to set some guidelines and standards about these issues, but sources of variability in experimentation are domain specific.

Next Steps - What Can We Do as a Community?

The final discussion session wrapped up the workshop by exploring the following steps and asking what the community can do to improve the sharing and reuse of cybersecurity research artifacts. David Balenson (USC Information Sciences Institute) facilitated the session and went around the virtual “room,” giving each participant a chance to share their thoughts and suggestions.

We briefly summarize the discussion for this session.

Create incentives for authors and evaluators; Authors need funding and recognition in the community for sharing usable artifacts. Evaluators need to receive more recognition too. Creating venues where artifacts can be published would help. Over time, these venues can become as prestigious as top security venues.

Build a culture of sharing and reuse: It is yet unclear how to do this, but we must recognize that it is a process, and we need to keep moving in this direction. Conferences promoting artifact sharing/reuse, review processes encouraging artifact submissions, and competitions all help. Teaching students early in their degree program that artifacts are essential and incorporating artifacts into classes also helps. The community needs to build incentives around artifact maintenance and take ownership of old but valuable artifacts, so they remain useful over time. This contribution could become a valued item on a student’s resume.

Provide funding for artifacts: Some supplemental funding to NSF projects with required quality checks. Also, NSF could provide funds to conferences for the AE process - this funding can compensate student evaluators and provide additional incentives.

Standardize processes: We need a consistent way to cite artifacts and permanently archive them somewhere. We also need to specify metadata that should be shared for each type of artifact and provide examples. We need guidelines for authors on how to prepare their artifacts for evaluation and sharing, and we need guidelines for evaluators. We also need guidelines around how to contribute to existing artifacts. We need guidelines on how to record and share research methods.

Artifact Survey

To gain input from a wider community of researchers, we constructed and distributed an artifact survey, which USC's Institutional Review Board (IRB) approved. The anonymous survey asked several questions about the participants' experiences with sharing and reusing artifacts. A copy of the survey is included in Appendix F.

In September 2022, we distributed the survey on mailing lists for networking and security researchers (geni-annouce@geni.net) and general researchers (itc@comsoc.org, mycolleagues@mailman.ufsc.br). We also posted the survey in NSF SaTC and NeTS researcher Slack groups. The survey is still available (https://forms.gle/FZN9LsY9h39suh2K7) for community members to take. Here we provide a preliminary summary of the survey responses.

By December 2022, the survey had received 31 responses. The survey asked the respondent to describe in their own words the area of research in which they work. We grouped responses into similar research field categories. Seventeen participants worked in networking/systems, six in cybersecurity, three in machine learning/natural-language processing, three in Mobile/IoT, and two in high-performance computing.

The survey further asked participants to state the length of time they worked on research, including graduate school. The response distribution was nearly even across 1-5 years, 5-10 years, 10-20 years, and more than 20 years.

The survey asked, "When you are starting a new research project, how often do you look for artifacts from other researchers?". 45% of respondents indicated that they always look for artifacts, 38% look often, and the rest look rarely or never. These responses indicate that **the demand for computational artifacts is high.**

The survey asked, "What percentage of the time are you successful in finding a possible candidate artifact (before evaluating it)?" Around 60% of respondents indicated they could find an artifact at least half the time, while the rest were less successful in finding candidate artifacts. More than a quarter of respondents said they could find a candidate artifact only 10% of the time. These answers indicate that **computational artifact findability is a significant problem for researchers.** The distribution of respondent areas of research and experience in the field are representative across all group areas and experience levels. This distribution suggests that the artifact findability problem is pervasive among different research fields and levels of research expertise.

The survey asked where participants searched for artifacts - this response allowed multiple options to be selected. More than 85% of participants utilized the Google search engine and

looked at published papers in the field to see if they referenced available artifacts. Github repository search was also popular (more than 75% of participants). Less popular approaches were asking colleagues in the field (51%), looking at Web sites of prominent researchers (25%), or using Zenodo (9%). Almost 16% of respondents indicated they searched “other” sources for artifacts. These other sources were primarily valuable for dataset or ML model search, such as ML model repositories, Ubuntu code repositories, AWS open data, etc.

The survey asked, “What percentage of the time are you able to use the artifact you found?” Answers varied widely here, with 70% of respondents indicating that they can use an artifact less than half the time. Almost 20% of respondents replied that they could use an artifact less than 10% of the time. These responses indicate that **computational artifact usability is a big issue for researchers.**

The survey asked respondents why their attempts to reuse artifacts have failed. We grouped and summarized responses from this free-response question. Almost all respondents indicated that code issues (missing dependencies, code not compiling, code depending on outdated packages or environments), data issues (too anonymized, too aggregated, too small, missing), and incomplete documentation (missing metadata, missing test scripts or README files) were the main factors for failure to reuse artifacts. Two responses also cited a flawed research methodology that impacted artifact quality. **These responses indicate that ensuring artifact quality and packaging is necessary to improve artifact usability.**

The survey asked respondents to estimate how long it takes them to learn how to use an artifact. Responses varied widely from a few days to a few weeks and sometimes even a few months. Most frequent answers ranged in 1-4 week period. **This result further highlights the need to improve artifact quality to facilitate reuse.**

The survey asked respondents for their opinion on what would improve artifact quality. We grouped and summarized responses from this free-response question. Respondents suggested: (1) more artifact evaluations with stricter standards and with requiring quality artifacts for paper publication and (2) metadata standards, test suites, and better documentation. Respondents also acknowledged that producing high-quality artifacts heavily depends on authors, as was reflected in suggestions for (3) close contact with authors to enable reuse and (4) increased funding for authors to release high-quality artifacts.

Since Docker has benefited code packaging and distribution, the survey asked respondents if Docker-packaged artifacts would work for their research. Around 60% indicated this would work for them, with the rest responding “rarely or never.” The main reasons for the inability to use Docker were the requirements of the research field (e.g., the research required distributed computation or complex installation, or the timing of operations would be jeopardized with containers), inability to interface with Docker (e.g., how to incorporate the Dockerized artifact into an existing, non-Dockerized workflow) and being unfamiliar with Docker. Several respondents pointed out that Dockerization only solves the “running” problem, while poor

assumptions, research methodology, and documentation remain to be solved. These responses indicate that **the use of Docker is beneficial to a great extent, and artifact authors should be encouraged to use it where possible.**

The survey asked respondents to provide other suggestions for packaging and sharing artifacts. The focus here was on the easy integration of code with researchers' workflows. Approaches ranged from sharing VM images, videos of how to run the code, tests, and documentation, sharing code as Python or Ubuntu packages (if applicable), and sharing platforms where code would "just run." Thus, **any actions to improve the ease of reuse would be beneficial.**

The survey asked respondents about their artifact-sharing practices to understand the needs of artifact authors. Almost 80% of respondents shared their research artifacts "always or often." The main reasons for sharing were altruism (helping other researchers), paying it forward, and the expected impact of sharing on one's research. On the other hand, the main reasons for not sharing were the overhead of sharing and being too early in the research process to share. These responses indicate that **researchers are generally willing to share their artifacts but may need additional support to make artifacts easy to reuse.**

The survey asked respondents what would motivate them to share more. The primary motivators were more professional recognition, feedback (to reward sharing), and more funding (to facilitate time investment).

The survey sought to understand how the research community practices sharing and reuse. It asked participants if it is common for researchers in their field to share and reuse artifacts. Almost 80% indicated that it was "very common or somewhat common" to share. Similarly, around 70% stated that compared to the prior work in publications via artifact reuse or reimplementation was "very common or somewhat common" in their field.

Key Findings

The workshop and survey together produced many findings around future needs and challenges for artifact sharing and reuse. We summarize here the key findings.

Artifact findability: There are currently multiple channels where researchers share their artifacts, such as Github repositories, lab and personal Web pages, Zenodo, Dropbox, etc. Some of these approaches to sharing may become obsolete when a project ends or a researcher moves from a given institution. There is a need to improve the ability to find artifacts by providing: (1) ways to index various shared artifacts on multiple platforms and provide a common repository for this index and (2) locations where artifacts can be permanently archived.

Artifact quality and usability: There is a great need to improve artifact quality and usability to save time for researchers looking to reuse artifacts. Some approaches the workshop and survey participants identified include better code packaging, better metadata, and better documentation for artifacts. To improve these elements, it is clear that the community needs to create some standards around packaging, metadata, and documentation and then teach young researchers how to produce artifacts meeting these standards.

Artifact maintenance: Many artifacts are released but not maintained, often due to the significant burden of ongoing maintenance, the unclear payoff for authors, and due to authors graduating or moving on to new projects. There is a need to come up with solutions to these challenges, although the community acknowledges that the problem is complex and that only partial solutions may be possible.

Incentives and funding: Artifact authors and evaluators need more recognition/value for their work. Ideally, there would be intrinsic incentives for artifact sharing and reuse that would work in concert with promotion and tenure pathways, career pathways, and funding. Authors need more funding to produce higher-quality artifacts since this involves significantly more work than that required to publish their research results.

Support for the entire research lifecycle: There is a need to support artifacts beyond datasets and code by capturing and sharing the whole research lifecycle. Things to be captured include research methods, such as hypothesis, workflow, experiment design, experimentation environment, assumptions, limitations, etc.

Conclusions and Next Steps

The overall goal of the artifacts workshop was to explore the issues and challenges researchers face in producing, evaluating, and reusing high-quality cybersecurity research artifacts, including datasets and software, as well as experiment designs, execution, and results.

During the workshop, attendees participated in a series of highly interactive sessions to discuss artifact evaluation; artifact repositories, indices, and challenges; artifact evaluation processes at conferences; artifact quality and metadata standards; going beyond datasets and code; and next steps - what can we do as a community? Additionally, we prepared and conducted an artifact survey to gain input from a wider community of researchers beyond the workshop participants. Together, the workshop discussions and survey results pointed toward several immediate next steps for the community.

Standardize requirements and processes around artifact sharing: A frequent topic of discussion was low artifact quality, artifacts getting obsolete over time, and artifacts not being portable. One prerequisite to improve on these issues is to standardize requirements and processes around artifact sharing and to create educational materials that teach young researchers how to produce compliant artifacts.

Create incentives for authors and evaluators: Today, there is much work for artifact authors and evaluators with low benefits. Authors currently benefit more from working on research to publish a new paper than from putting a similar amount of time into preparing, releasing, and maintaining an artifact associated with an existing publication. Evaluators spend much time on artifact evaluation but only receive name recognition as part of the evaluation committee. In both cases, we should create professional and monetary incentives for authors and evaluators to offer a higher payoff for artifact sharing and evaluation. Some discussions touched on the possibility of requiring artifacts to be published when a research publication is accepted. Recently many venues have started encouraging artifact publication. A few venues require artifact submission (e.g., ACM CCS, Journal of Systems Research).

Build a culture of sharing and reuse: Conferences and journals can help build a culture of sharing and reuse in the research community through continued calls for artifact publication and encouragement that the research described in paper submissions be evaluated against published artifacts. Further, community guidelines, tutorials, university classes, and software packages that help teach young researchers how to share and reuse artifacts will help build a culture of sharing and reuse.

Provide funding for artifact sharing: The main hurdle for artifact authors to provide higher-quality artifacts is the time and resources required for this task. Funding agencies should consider supplemental funding vehicles to help fund artifact-sharing efforts with clear artifact measurement requirements. Such funding vehicles should not be too competitive or need

significant funding. Support for 3-6 months of a graduate student should suffice to produce a high-quality artifact in many cases. This level of funding is comparable to other supplemental funding, e.g., for REUs.

Provide research infrastructure for artifacts sharing: Limited access to resources for sharing and reusing high-quality artifacts, especially in specialized environments or on specialized equipment, also limits sharing and reuse. The community would benefit significantly from robust online catalogs of research artifacts, such as SEARCCH [SEARCCH-HUB], with rich metadata and information about artifact quality and applicability. The community would also benefit from an experimentation environment with full support for the research lifecycle, including artifact packaging for sharing and reuse, automated execution scripts, ample data storage, and experiment state saving and restoration. Artifact evaluation committees could use such infrastructure to offer evaluation services directly on the infrastructure. Once the evaluation completes, its records, evaluator notes, and packaged experiments could remain housed on the infrastructure to be easily reused by others in the same environment.

The robust workshop and survey participation, wide-ranging perspectives, and productive discussion sessions and comments demonstrated a strong need for a coordinated effort to improve the quality and reuse of cybersecurity research artifacts. The workshop was the first in what we hope to be an ongoing series of discussions and interactions and the start of a new, vibrant community centered around the sharing and reusing of high-quality cybersecurity research artifacts. We plan to hold additional workshops, likely on an annual basis. We will target holding the next workshop in the fall of 2023 and hope to learn about new efforts and initiatives relating to improving the quality and reuse of artifacts.

Acknowledgements

The workshop organizers thank all participants who took the time to attend the workshop and for their open and collegial participation.

The workshop was supported by the Sharing Expertise and Artifacts for Reuse through Cybersecurity Community Hub (SEARCCH) project which is supported by the National Science Foundation under grant numbers 1925773, 1925616, 1925588, 1925564. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[ACM-ARTIFACTS] ACM Artifact Review and Badging, Version 1.1, August 24, 2020. https://www.acm.org/publications/policies/artifact-review-and-badging-current

[ACSAC-AE17] Artifact Submission, Call for Submissions, 2017 Annual Computer Security Applications Conference (ACSAC) (website). https://www.acsac.org/2017/submissions/papers/artifacts/

[ACSAC-AE22] Paper Artifacts, 2022 Annual Computer Security Applications Conference (ACSAC) (website). https://www.acsac.org/2022/submissions/papers/artifacts/

[ACSAC-COMP] Cybersecurity Artifacts Competition and Impact Award, Annual Computer Security Applications Conference (ACSAC). https://www.acsac.org/2022/submissions/artifacts_competition/

[BALENSON] David Balenson, Terry Benzel, Eric Eide, David Emmerich, David Johnson, Jelena Mirkovic, and Laura Tinnel. 2022. Toward Findable, Accessible, Interoperable, and Reusable Cybersecurity Artifacts. In Proceedings of the 15th Workshop on Cyber Security Experimentation and Test (CSET '22). Association for Computing Machinery, New York, NY, USA, 65–70. https://doi.org/10.1145/3546096.3546104

[BAROWY] Dan Barowy, Charlie Curtsinger, Emma Tosch, John Vilk, HOWTO for AEC Submitters, May 2020. http://bit.ly/HOWTO-AEC

[BELLER] Mirotz Beller, Why I will never join an Artifacts Evaluation Committee Again, June 26, 2020. https://inventitech.com/blog/why-i-will-never-review-artifacts-again/

[CCS-AE22] Providing Artifacts, Call for Papers, ACM Conference on Computer and Communications Security (CCS) (website). https://www.sigsac.org/ccs/CCS2022/call-for/call-for-papers.html

[CEF] Cybersecurity Experimentation of the Future (CEF) (website). https://cef.cyberexperimentation.org/.

[CEF-REPORT] David Balenson, Laura Tinnel, and Terry Benzel, Cybersecurity Experimentation of the Future: Catalyzing A New Generation of Experimental Cybersecurity Research, Final Report, July 31, 2015. https://cef.cyberexperimentation.org/application/files/2616/2160/7871/CEF_Final_Report_Bound_20150922.pdf.

[FAIR] FAIR Principles (website). https://www.go-fair.org/fair-principles/

[MIRKOVIC] Jelena Mirkovic, Genevieve Bartlett, and Jim Blythe, DEW: Distributed Experiment Workflows, 11th USENIX Workshop on Cyber Security

Experimentation and Test (CSET 18), August 2018, Baltimore, MD. https://www.usenix.org/conference/cset18/presentation/mirkovic.

[MYTKOWICZ] Todd Mytkowicz, Amer Diwan, Matthias Hauswirth, and Peter F. Sweeney. 2009. Producing wrong data without doing anything obviously wrong! SIGPLAN Not. 44, 3 (March 2009), 265–276. https://doi.org/10.1145/1508284.1508275

[NDSS] Network and Distributed System Security (NDSS) Symposium (website). https://www.ndss-symposium.org/

[PADHYE] Rohan Padhye, Artifact Evaluation: Tips for Authors, August 7, 2019. https://blog.padhye.org/Artifact-Evaluation-Tips-for-Authors/

[SEARCCH] Sharing Expertise and Artifacts for Reuse through a Cybersecurity Community Hub (website). https://searcch.cyberexperimentation.org/

[SEARCCH-HUB] SEARCCH Hub (website). https://hub.cyberexperimentation.org/

[SECARTIFACTS] Security Research Artifacts (website). https://secartifacts.github.io/

[USENIX] USENIX Security Symposia (website). https://www.usenix.org/conferences/byname/108

[USENIX-AE22] USENIX Security '22 Call for Artifacts (website). https://www.usenix.org/conference/usenixsecurity22/call-for-artifacts

[USENIX-AE22-GUIDE] USENIX Security '22 Artifact Appendix Guidelines (V20220119) (website). https://www.usenix.org/conference/usenixsecurity22/artifact-appendix-guidelines

[WILKINSON] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[XU] Tianyin Xu, How Are Award-winning Systems Research Artifacts Prepared (Part 1), January 8, 2021. https://www.sigops.org/2021/how-are-award-winning-systems-research-artifacts-prepared-part-1/

Appendices

Appendix A: List of Acronyms

ACM	Association for Computing Machinery
ACSAC	Annual Computer Security Applications Conference
AE	Artifact Evaluation
AEC	Artifacts Evaluation Committee
AWS	Amazon Web Services
CCRI	CISE Community Research Infrastructure
CEF	Cybersecurity Experimentation of the Future
CFP	Call for Papers
FAIR	Findable, Accessible, Interoperable, and Reusable
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
ML	Machine Learning
NDSS	Network and Distributed System Security Symposium
NeTS	Networking Technology and Systems
NIST	National Institute of Standards and Technology
NSF	National Science Foundation
PETS	Privacy Enhancing Technologies Symposium
REU	Research Experiences for Undergraduates
SATC	Secure and Trustworthy Cyberspace
SEARCCH	Sharing Expertise and Artifacts for Reuse through a Cybersecurity Community Hub
TTP	Transition to Practice
URL	Uniform Resource Locator
VM	Virtual Machine

Appendix B: Workshop Invitation

From: David Balenson
To: David Balenson
Bcc: [Invitees]
Cc: Terry Benzel, Jelena Mirkovic, Laura Tinnel, Eric Eide, David Johnson, David Emmerich
Subject: INVITATION: Virtual Workshop on Cybersecurity Research Artifacts

Dear Colleague,

Researchers in experimental cybersecurity are increasingly sharing the code, data, and other artifacts associated with their studies. This trend is encouraged and rewarded by conferences and journals through practices such as artifact evaluation and badging. While these trends in sharing artifacts are promising, the cybersecurity community is still far from an ecosystem in which artifacts are FAIR: findable, accessible, interoperable, and reusable [1]. The lack of established standards and best practices for sharing and reuse results in artifacts that are often difficult to find and reuse; in addition, the lack of community standards results in artifacts that may be incomplete and low-quality.

As a researcher in the community who is producing, evaluating, and/or using experimental artifacts in your research, we invite you to participate in a virtual workshop, **Improving the Quality and Reuse of Cybersecurity Datasets, Software, and Other Research Artifacts**, to be held on **Thursday, September 15, 2022 from 10am-4pm EDT (UTC-4)**.

The goal of the workshop is to explore the issues and challenges researchers face in producing, evaluating, and reusing high-quality cybersecurity research artifacts, including datasets and software, as well as experiment designs, execution, and results. The workshop will bring together researchers like yourself who are producing, evaluating, and reusing experimental artifacts. The agenda will consist of a series of highly interactive sessions in which participants will be invited to share their thoughts on how to package and describe artifacts, including their intended uses and limitations; and how to better share and find relevant artifacts, including knowledge and experience about their use.

Some of the topics we plan to discuss and explore include:

- Reflections on artifact evaluation
- SEARCCH and other artifact repositories
- Artifact evaluation processes at conference
- Artifact quality and metadata standards
- Beyond datasets and code
- Next steps - what can we do as a community?

Please come prepared to share your experiences and ideas as part of these facilitated discussions. The ultimate goal is to learn more about potential approaches and opportunities to increase the effectiveness of experimental artifacts and broaden their reuse.

Please RSVP to David Balenson <david.balenson@sri.com> to let us know if you and/or others from your team can participate. Additional information, including a detailed agenda and Zoom link, will be distributed prior to the workshop.

We greatly appreciate your consideration of this invitation and look forward to your participation in the workshop!

Sincerely,
David Balenson (SRI International)

On behalf of the workshop team:
Terry Benzel, Jelena Mirkovic (USC-ISI)
Laura Tinnel, David Balenson (SRI International)
Eric Eide, David Johnson (U. Utah)
David Emmerich (U. Illinois)

[1] David Balenson, Terry Benzel, Eric Eide, David Emmerich, David Johnson, Jelena Mirkovic, and Laura Tinnel. 2022. Toward Findable, Accessible, Interoperable, and Reusable Cybersecurity Artifacts. In Proceedings of the 15th Workshop on Cyber Security Experimentation and Test (CSET '22). Association for Computing Machinery, New York, NY, USA, 65–70. https://doi.org/10.1145/3546096.3546104

ADDITIONAL BACKGROUND INFORMATION

The cybersecurity R&D community needs to share cybersecurity research artifacts, including datasets and software, as well as experiment designs, execution, and results, in a way that lowers the barrier to sharing and reuse. The evaluation of the cybersecurity properties of computer, networking, and cyber-physical research is frequently performed in ad hoc ways, which severely hinders scientific progress. Many researchers use a combination of methods and infrastructure to conduct experiments using one-off, painstaking, and error-prone processes with limited sharing for reuse and validation. The lack of repeatable, reproducible, and reusable processes and other artifacts limits the ability to build on the work of others or to compare solutions. Enabling the rapid sharing and reuse of experiment artifacts is crucial to improving our rate of progress and will help transform our scientific community.

A number of cybersecurity conferences and workshops, including ACSAC and Usenix Security, have artifact initiatives that encourage authors of accepted papers to submit software and data artifacts and make them publicly available to the entire community. Artifacts are evaluated and authors of verified artifacts are rewarded with an ACM Artifacts Evaluated badge on their

papers. Various platforms, including GitHub, Zenodo, and the NSF-funded SEARCCH hub, are available to help researchers share artifacts they produce or find artifacts produced by other researchers. But, is the cybersecurity research community taking full advantage of these initiatives and capabilities? If not, why not? What can we as researchers do to lower the barrier and improve sharing? And, what else can be done to incentivize sharing? Furthermore, while the sharing of code and data artifacts is increasingly understood, less is known or understood about how to capture, share, and archive entire experiments. How can we capture, share, and archive experiment designs, execution, and results? What is reproducible and what is implementation specific?

This workshop will explore the issues and challenges researchers face in capturing, packaging, sharing, finding, and reusing cybersecurity research artifacts, including datasets and software, as well as experiment designs, execution, and results. Participants will discuss and share experiences regarding current evaluation initiatives, sharing infrastructure, and incentives for sharing and reusing artifacts. We will also explore and identify other potential approaches and opportunities to increase and improve the sharing of experimentation artifacts as a solid practice in the cybersecurity research community.

Appendix C: Workshop Agenda

Improving the Quality and Reuse of Cybersecurity Datasets, Software, and Other Research Artifacts

VIRTUAL WORKSHOP

September 15, 2022

Google Drive Folder:
https://drive.google.com/drive/folders/1J5PF86ZTu2O3pPV1gPjY-p82bl56PwDp?usp=sharing

AGENDA

Thursday, September 15, 2022 – all times EDT (UTC-4)	
09:45-10:00	**GATHERING**
10:00-10:30	**Welcome, Introductions, and Background**
10:30-11:20	**DISCUSSION: Reflections on Artifact Evaluation** In the past decade, artifact evaluation has become a feature of many conferences in cybersecurity and other areas of computer science. Widespread practice is for artifacts to be solicited from the authors of accepted papers. Papers receive badges if their corresponding artifacts are judged to be available, functional, reusable, and/or useful for reproducing the papers' main results. ● How has the adoption of artifact evaluation affected your practices and/or expectations of published research? ● How well is the current adoption and practice of artifact evaluation serving the community? ● What has the community learned from the current practice of artifact evaluation? ● In what ways should current practice evolve to better serve the research community? **Facilitator**: Eric Eide, University of Utah **Notes**: https://docs.google.com/document/d/1CfVSimzFy55TQvVjtEIKfJzGgOJUMHQM/edit?usp=sharing&ouid=114687916029723919981&rtpof=true&sd=true

11:20-11:30	**BREAK**
11:30-12:20	**DISCUSSION: Artifact Repositories, Indices, and Challenges** Discuss artifact repositories such as GitHub and Zenodo, indices such as FindResearch.org and SEARCCH, and the current state of artifact packages that impede index performance and reuse. ● How do we balance the need for sharing with "useful" sharing? ● How much effort is reasonable for initially adding artifacts to indices? ● Is crowdsourcing artifact use knowledge (over time) helpful to researchers? ● What additional catalog features would be the most useful to the cybersecurity research community? **Facilitator**: Laura Tinnel, SRI International **Notes**: https://docs.google.com/document/d/1FU-uwdxAH9XW1i1zbrA7lbo_5vopU5EX/edit?usp=sharing&ouid=114687916029723919981&rtpof=true&sd=true
12:20-12:30	**BREAK**
12:30-13:20	**DISCUSSION: Artifact Evaluation Initiatives at Conferences** Discuss the conference artifacts evaluation initiatives and processes at events such as ACSAC, USENIX Security, and PETS. ● What works well with the current artifact evaluation initiatives? ● What guidance is provided to artifact submitters? To artifact evaluators? ● How could the current processes be improved? ● How could the community capture the knowledge gained during the artifacts evaluation process? **Facilitator**: David Balenson, USC-ISI **Notes**: https://docs.google.com/document/d/1mO9EiMoc-EwcSD9xOf-bAUJWmEaiOcNI/edit?usp=sharing&ouid=114687916029723919981&rtpof=true&sd=true
13:20-13:30	**BREAK**
13:30-14:20	**DISCUSSION: Artifact Quality and Metadata Standards** An often mentioned obstacle to artifact reuse is artifact quality. Researchers do not like to spend a lot of time on an unknown artifact, just to discover it will not meet their needs. ● How would you define artifact quality? Is it something generic or something specific to a research goal? ● What steps can authors take to ensure their artifacts are of high quality? ● How can a researcher evaluate the quality of an artifact they consider reusing? ● Are some artifact packaging approaches (e.g., Docker, VM) easier to reuse?

	• How can we motivate crowdsourcing of artifact quality evaluation? **Facilitator**: Jelena Mirkovic, USC-ISI **Notes**: https://docs.google.com/document/d/119XFbIae6rwpfhn1lDeI_m12JeDOKHX5/edit?usp=sharing&ouid=114687916029723919981&rtpof=true&sd=true
14:20-14:30	**BREAK**
14:30-15:20	**DISCUSSION: Beyond Datasets and Code** There has been a rapid increase in sharing of data and code through a wide range of processes, infrastructure and reward mechanisms. The vast majority of shared artifacts are code and data tied to a publication. What are the other artifacts of research that should be shared? • Can we share experimental research artifacts? • How do we capture and share experiment designs, execution, and results? • How much information is needed about research infrastructure to share an "experiment"? • How do we share process, steps, and life-cycle meaningfully? **Facilitator**: Terry Benzel, USC-ISI **Notes**: https://docs.google.com/document/d/1kej0yDveJ0mk_2CsEvu9yh2gK9AT-Z_D/edit?usp=sharing&ouid=114687916029723919981&rtpof=true&sd=true
15:20-15:50	**NEXT STEPS: What Can We Do as a Community?** Discuss what steps we as a community can take to further improve the quality and reuse of cybersecurity datasets, software, and other research artifacts **Facilitator**: David Balenson, USC-ISI **Notes:** https://docs.google.com/document/d/1c2JHukTfr2Fvsw9BA7fJ01J3Eyv0J0ij/edit?usp=sharing&ouid=114687916029723919981&rtpof=true&sd=true
15:50-16:00	**Wrap-up**
16:00	**ADJOURN**

Appendix D: Workshop Presentations

Introduction and Background - David Balenson (USC-ISI)
Slides:
https://docs.google.com/presentation/d/1kvm9WKxDgFZp3HhIvF95AjUllBh4dk14/edit?usp=sharing&ouid=114687916029723919981&rtpof=true&sd=true

Reflections on Artifact Evaluation - Eric Eide (IUtah)
Slides: https://docs.google.com/presentation/d/1vz6-tinBJLTMjVKHdOziPmwUnXMRkKwS/edit?usp=sharing&ouid=114687916029723919981&rtpof=true&sd=true

Artifact Repositories, Indices, and Challenges - Laura Tinnel (SRI International)
Slides:
https://docs.google.com/presentation/d/1zPih2n1exz1mXzM0a5u73fzqRtN01kmn/edit?usp=sharing&ouid=114687916029723919981&rtpof=true&sd=true

Artifact Evaluation Initiatives at Conferences - David Balenson (USC-ISI)
Slides:
https://docs.google.com/presentation/d/1PgFRuN0gACkdkFsoJb8C47LsOLtkwMLF/edit?usp=sharing&ouid=114687916029723919981&rtpof=true&sd=true

Beyond Datasets and Code - Terry Benzel (USC-ISI)
Slides:
https://docs.google.com/presentation/d/1sfKxgTIoxxlMMdsLndC_9J4xG55xACKE/edit?usp=sharing&ouid=114687916029723919981&rtpof=true&sd=true

Appendix E: Summary of Artifact Evaluation Initiatives at Cybersecurity Conferences

The following table summarizes the artifact evaluation initiatives at ACSAC, USENIX Security, and ACM CCS through 2023.

Year	Initiative (Submission & Evaluation)	Stats	Badges	AEC
ACSAC 2017 https://www.acsac.org/2017/, https://www.acsac.org/2017/artifacts/	Encouraged authors of accepted papers to submit artifacts; evaluated after paper accepted	12 out of 48 accepted papers with evaluated artifacts (25%)	ACM Artifact Evaluated | Functional	Manuel Egele, 15 members
ACSAC 2018 https://www.acsac.org/2018/, https://www.acsac.org/2018/artifacts/	Encouraged authors of accepted papers to submit artifacts; evaluated after paper accepted	22 out of 60 accepted papers with evaluated artifacts (37%)	ACM Artifact Evaluated | Functional	Manuel Egele, 15 members
ACSAC 2019 https://www.acsac.org/2019/submissions/papers/artifacts/, https://www.acsac.org/2019/program/artifacts/	Encouraged authors of accepted papers to submit artifacts; evaluated after paper accepted	20 out of 60 accepted papers with evaluated artifacts (33%); 12 functional 8 reusable	ACM Artifact Evaluated | Functional ACM Artifact Evaluated | Reusable	Roberto Perdisci, 54 members
ACSAC 2020 https://www.acsac.org/2020/submissions/papers/artifacts/, https://www.acsac.org/2020/program/artifacts/	Encouraged authors of accepted papers to submit artifacts; evaluated after paper accepted	26 out of 70 accepted papers with evaluated artifacts (37%); 16 functional 10 reusable	ACM Artifact Evaluated | Functional v1.1 ACM Artifact Evaluated | Reusable v1.1	Roberto Perdisci, 50 students and 27 mentors
ACSAC 2021 https://www.acsac.org/2021/submissions/papers/artifacts/, https://www.acsac.org/2021/program/artifacts/	Encouraged authors of accepted papers to submit artifacts; early ad-hoc artifact evaluation by chairs factored into Round 2 paper decision; traditional evaluation after paper accepted	20 out of 80 accepted papers with evaluated artifacts (25%); 11 functional 9 reusable	ACM Artifact Evaluated | Functional v1.1 ACM Artifact Evaluated | Reusable v1.1	Martina Lindorfer and Gianluca Stringhini, 39 students and 18 mentors

Year	Initiative (Submission & Evaluation)	Stats	Badges	AEC
ACSAC 2022 https://www.acsac.org/2022/submissions/papers/artifacts/, https://www.acsac.org/2022/program/artifacts/	Encouraged authors of accepted papers to submit artifacts; early ad-hoc artifact evaluation by chairs factored into Round 2 paper decision; traditional evaluation after paper accepted	43 out of 73 accepted papers with evaluated artifacts (59%); • 25 functional • 18 reusable	ACM Artifact Evaluated | Functional v1.1 ACM Artifact Evaluated | Reusable v1.1	Martina Lindorfer and Gianluca Stringhini, 57 students 22 mentors
ACSAC 2023 https://www.acsac.org/2023/submissions/papers/artifacts/	Encouraged authors of accepted papers to submit artifacts; early ad-hoc artifact evaluation by chairs factored into Round 2 paper decision; traditional evaluation after paper accepted	TBD	ACM Artifact Evaluated | Functional v1.1 ACM Artifact Evaluated | Reusable v1.1	Adwait Nadkarni and Xiaojing Liao
USENIX Security 2020 https://www.usenix.org/conference/usenixsecurity20/call-for-artifacts	Optional submission and evaluation after paper (conditionally) accepted	23 papers with badges	USENIX Artifact Evaluated - Passed	Thorsten Holz and Brendan Dolan-Gavitt, 35 members
USENIX Security 2021 https://www.usenix.org/conference/usenixsecurity21/call-for-artifacts	Optional submission and evaluation after paper (conditionally) accepted	34 papers with badges	USENIX Artifact Evaluated - Passed	
USENIX Security 2022 https://www.usenix.org/conference/usenixsecurity22/call-for-artifacts	Encouraged optional submission and evaluation after paper (conditionally) accepted; Badges appear on Artifact Appendix; Papers can be evaluated against multiple badges	114 papers with badges; • 107 Artifact Available • 98 Artifact Functional • 65 Results Reproduced	USENIX Artifact Evaluated - Artifact Available USENIX Artifact Evaluated - Artifact Functional USENIX Artifact Evaluated - Results Reproduced	Clémentine Maurice and Cristiano Giuffrida, 104 members

Year	Initiative (Submission & Evaluation)	Stats	Badges	AEC
USENIX Security 2023 https://www.usenix.org/conference/usenixsecurity23/call-for-artifacts	Encouraged optional submission and evaluation after paper (conditionally) accepted; Badges appear on Artifact Appendix; Papers can be evaluated against multiple badges	Summer Cycle: 20 out of 83 accepted papers (24%) • 19 Artifact Available • 18 Artifact Functional • 13 Results Reproduced	USENIX Artifact Evaluated - Artifact Available USENIX Artifact Evaluated - Artifact Functional USENIX Artifact Evaluated - Results Reproduced	Cristiano Giuffrida and Anjo Vahldiek-Oberwagner, 102 members
ACM CCS 2022 https://www.sigsac.org/ccs/CCS2022/call-for/call-for-papers.html	Submissions whose claimed contributions rely on artifacts are expected to make these accessible to the reviewers, unless there are good reasons not to, in which case these reasons must be mentioned in the submission. Submissions whose claimed contributions do not rely on artifacts do not need to submit artifacts. No artifact evaluation.	N/A	N/A	N/A

Year	Initiative (Submission & Evaluation)	Stats	Badges	AEC
ACM CCS 2023 https://www.sigsac.org/ccs/CCS2023/call-for-paper.html	Submissions whose claimed contributions rely on artifacts are expected to make these accessible to the reviewers, unless there are good reasons not to, in which case these reasons must be mentioned in the submission. Submissions whose claimed contributions do not rely on artifacts do not need to submit artifacts. Encouraged optional submission for evaluation after paper accepted.	TBD	TBD	TBD

Appendix F: Surveying Sharing and Reuse of Computational Artifacts

Surveying Sharing and Reuse of Computational Artifacts
Principal Investigator: **Jelena Mirkovic, USC Information Sciences Institute**

You are invited to participate in a research study. Your participation is voluntary. This document explains information about this study. You should ask questions about anything that is unclear to you.

PURPOSE

The purpose of this study is to identify common obstacles to sharing and reuse of computational artifacts. We hope to learn which obstacles exist in a given research field, how common they are, and any suggestions respondents have for overcoming the obstacles. You are invited as a possible participant because you are a researcher in scientific field and you may share or reuse computational artifacts.

PARTICIPANT INVOLVEMENT

You will be asked to respond to 20 short questions (multiple choice or one-sentence response) about your personal experiences around computational artifact sharing and reuse. It will take about 5 minutes to complete the survey.

If you decide to take part, you will be asked to complete an online survey.

CONFIDENTIALITY

The members of the research team, and the University of Southern California Institutional Review Board (IRB) may access the data. The IRB reviews and monitors research studies to protect the rights and welfare of research subjects. Your participation is anonymous. We do not collect any private identifying data in the study.

INVESTIGATOR CONTACT INFORMATION

If you have any questions about this study, please contact Jelena Mirkovic at mirkovic@isi.edu or at 310-448-9170.

IRB CONTACT INFORMATION

If you have any questions about your rights as a research participant, please contact the University of Southern California Institutional Review Board at (323) 442-0114 or email irb@usc.edu.

If you agree to participate in this study please continue, otherwise close this window. Thank you for your participation!

COMPUTATIONAL ARTIFACT SURVEY

This survey asks about computational artifact sharing and reuse in your field of research. A computational artifact is code or data (or a bundle containing multiple pieces of code and/or data), meant to be used for research. Please answer the following questions using your best recollection. Estimates are fine, you do not need to be very accurate.

1. What is your research field (e.g., cybersecurity, networking, etc)

2. How many years have you been working in your field of research (include graduate studies in this count)?

◯ Less than 1 year
◯ 1-5 years
◯ 6-10 years
◯ 10-20 years
◯ more than 20 years

3. When you are starting a new research project, how often do you look for artifacts from other researchers?

◯ Always
◯ Often
◯ Rarely
◯ Never

4. What percentage of time are you successful in finding a possible candidate artifact (before evaluating it)?

◯ less than 10%
◯ 10-25%
◯ 25-50%
◯ 50-75%
◯ 75-100%

5. When starting a new research project where do you look for artifacts you may reuse? Please select all that apply

- Published papers in the field
- Famous researcher Web pages
- Github repository search
- Google search / Google scholar search
- Ask colleagues in the field
- Zenodo
- Other

6. If you specified "other" in the previous question please provide more details here

7. What percentage of time are you actually able to use the artifact you found? (100% means that every artifact you find you can use)

- ◯ less than 10%
- ◯ 10-25%
- ◯ 25-50%
- ◯ 50-75%
- ◯ 75-100%

8. In your experience, what are some reasons that your attempts to reuse artifacts have failed in the past

9. In your opinion, what would help improve artifact quality?

10. How long (weeks, months?) does it take you on an average to learn how to use an artifact and make it work?

11. Would an artifact shared as a Docker image or a hypervisor image work for you?

- ◯ Always
- ◯ Often
- ◯ Rarely
- ◯ Never

12. If you have answered rarely or never to the question 11, please help us understand why

13. Please provide any suggestion you may have on how to share artifacts in a way that they are easier to reuse (via Docker, VM images, shared platforms or any other suggestion you may have)

14. Do you yourself share computational artifacts from your research?

◯ Always
◯ Often
◯ Rarely
◯ Never

15. If you have answered rarely or never to the question 14, please help us understand what prevents you from sharing more

16. If you have answered often or always to question 14, what is your motivation for sharing?

17. What would motivate you to share more than you already do?

18. Is it common in your research field for researchers to share artifacts?

◯ Very common
◯ Somewhat common
◯ Somewhat uncommon
◯ Very uncommon
◯ Unable to answer

19. Is it common in your research field to directly compare your research solution with published work using shared artifacts or reimplementing prior work?

◯ Very common
◯ Somewhat common
◯ Somewhat uncommon
◯ Very uncommon
◯ Unable to answer

20. Do you have any advice or opinion about how to improve artifact sharing and reuse in your research field?